Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

RECOMBINATION OF VIRUSES

Jozef J Bujarski, Northern Illinois University, Plant Molecular Biology Center, Montgomery Hall, DeKalb, Illinois, USA

Introduction

Biochemically, recombination is a process of creating new genomic molecules by combining or substituting pieces of nucleic acids. Genetically, recombination could be defined as physical exchange of fragments among the parental genetic material. The results of recombination are progeny genomes that contain genetic information in nonparental combinations. Recombination was recognized as an important factor producing the genetic diversity upon which natural selection can operate.

Recombination events can occur in both RNA and DNA viruses. Since the molecular events behind DNA and RNA recombination differ in many aspects, they are described separately below.

Recombination in DNA Viruses

Recombination in many DNA viruses is believed to be accomplished by cellular enzymatic activities. There are two general types of genetic DNA recombination in the cell: homologous recombination (general recombination) and nonhomologous recombination. Nonhomologous or site-specific recombination occurs relatively rarely and requires special proteins that recognize specific DNA sequences to promote recombination. Homologous recombination occurs between two DNA sequences that are the same or very similar in the region of crossovers. Homologous recombination probably occurs in every DNA-based organism and it happens much more often than nonhomologous recombination.

There is plenty of information on the biochemical pathways responsible for crossovers in DNA. Besides sequence identity the requirements of general recombination include complementary base-pairing between double-stranded DNA molecules, recombination enzymes and the formation of heteroduplex within the regions of complementary base-pairing between the two recombining DNA molecules. Studies of the enzymology of DNA recombination in bacteria (and in particular of RecA, RecBCD proteins of *Escherichia coli*) have led to a large amount of literature, including many general reviews. Related recombination activities have been found and studied in eucaryotic sources, including yeast, insect, mammalian and plant cells.

Some DNA viruses encode their own proteins that function during recombination processes. In fact some DNA viruses serve as model systems for the study of recombination. For instance, certain bacteriophages encode the recombination pathways in order to avoid dependence on host systems. Such recombination can be used for repairing damaged phage DNA and for exchanging DNA between related phages to increase their diversity. High-frequency illegitimate recombination was observed at the replication origin of bacteriophage M13 in the *E. coli* host. The crossovers occurred at the nucleotide adjacent to the nick at the replication origin, by joining to a nucleotide elsewhere in the genome. This implied a breakage-and-reunion mechanism of illegitimate recombination, operating in *E. coli*.

Many of the phage recombination activities are analogous to those present in the host bacteria and studies on bacterial recombination systems were influenced by studies on phage systems. These pathways include Rec proteins of phages T4 and T7 (analogous to host RecA, RecG, RuvC or RecBCD proteins), RecE pathway in the rac prophage of *E. coli* K-12, or the phage 1 *red* system. In phage lambda there is another recombination system that can substitute for the RecF pathway components in *E. coli*. Models illustrating functions of RecA protein, RecBCD enzyme and Ruv proteins are shown in **Fig. 1**. A cartoon of the correlation of different stages of DNA recombination with transcription and DNA replication during bacteriophage T4 growth cycle is shown in **Fig. 2**.

Analogies between selected phage and *E. coli* host recombination functions are shown in **Table 1**.

Recombination between viral DNA and host genes was first observed in transducing bacteriophages in procaryotes and for retroviruses in eucaryotes. In some cases this represents a useful way for DNA viruses to acquire cellular genes. Among interesting examples of acquisition of cellular genes by DNA viruses are tRNA genes present in bacteriophage T4. These genes contain introns indicating that bacteriophage T4 must have passed through a eucaryotic host during evolution.

Genetic methods that rely on the use of mutants have been one of the most popular approaches for studying genetic recombination of DNA viruses. In viruses that have a single-component DNA genome,

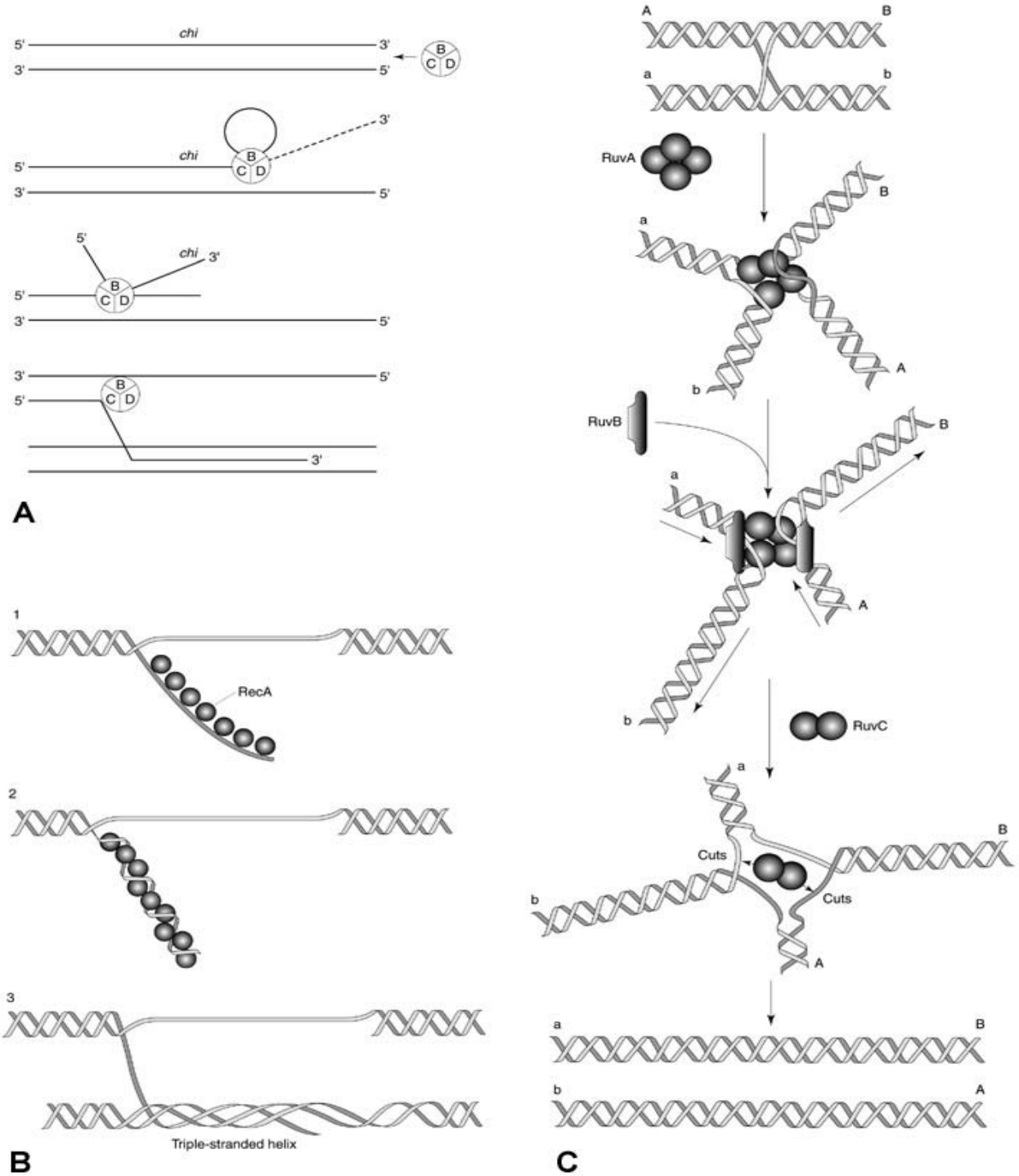

Figure 1 (**A**) Model for promotion of recombination initiation at a *chi* site by the RecBCD enzyme. The RecBCD enzyme loads on to the DNA at a free end or at a double-strand break internal to the DNA. It then moves along the DNA, displaces a loop, and the 3′ end is degraded by the exonuclease activity of the RecBCD enzyme (dotted line). The exonuclease activity is inhibited at a *chi* site, the 3′ end is no longer degraded and can thus invade another DNA molecule. (**B**) Synapse formation between two homologous DNAs by RecA protein. In steps 1 and 2, the RecA protein binds to the single-stranded end and forces it into an extended helical structure. In step 3, the helical single-stranded DNA can pair with a homologous double-stranded DNA in its major groove to form a stable extended triple helix. (**C**) Model for the mechanism of action of the Ruv proteins. RuvA binds to the Holliday junction. Note that the figure starts with one turn of blue-gray heteroduplex. Then RuvB binds to RuvA, and the junction migrates, deriving energy from ATP cleavage. RuvC cleaves two strands of the Holliday junction to resolve the junction into separate DNA molecules. Note the three turns of heteroduplex after junction migration. (Reproduced with permission from *Molecular Genetics of Bacteria* (1997) ASM Press.)

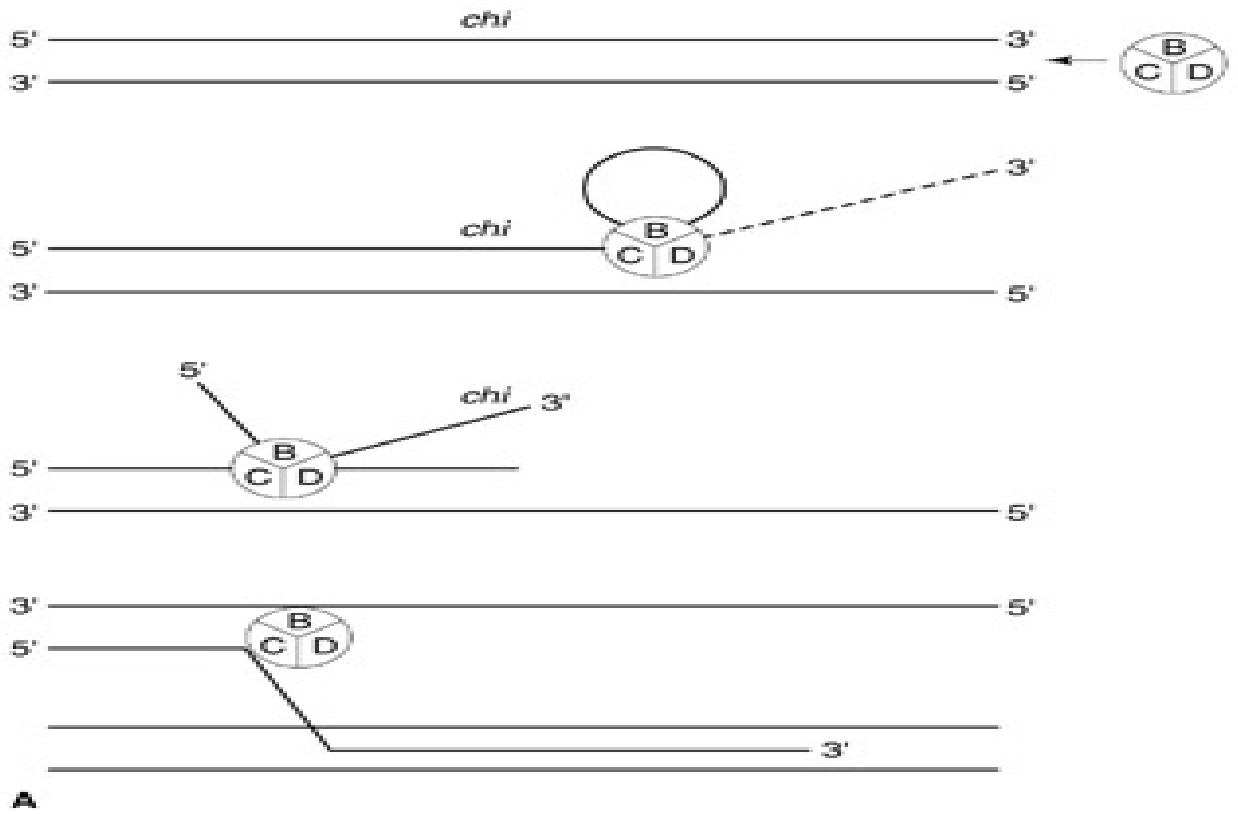

Figure 2 Cartoon showing the correlation between T4 recombination, and transcription and replication processes during bacteriophage T4 life cycle. The upper panel shows early (a), middle (b), and late (c) promoters recognized by forms of the host RNA polymerase that is modified by ADP-ribosylation of the α subunits and by T4-encoded accessory proteins. (d) shows a ribosome-binding site (RBS) of the mRNA for a late protein (Endo VII) that can be sequestered in a long early transcript but is free to initiate protein synthesis from a late transcript. The lower panel illustrates different stages of DNA replication and recombination in the context of the transcriptional program. (e) represents two T4 DNA molecules infecting a bacterium. (f) depicts bidirectional origin of replication cycle. (g) shows a replication–recombination pathway that becomes essential when the RNA polymerase modifications prevent further origin initiation. It is hypothesized that late after infection proteins required to initiate Okazaki pieces become limiting and that some join-cut-copy recombination requiring endonuclease VII occurs (h and i). Packaging proteins finally compete with replication and early recombination proteins for the recombinational intermediates, the concatamers are debranched, cut to headfull lengths, and packaged. (Reproduced with permission from *Molecular Biology of Bacteriophage T4* (1994) ASM Press.)

recombination can involve the exchange of DNA fragments. In contrast, in segmented DNA viruses, in addition to exchanges within each DNA component, the genome segments can be reassorted. This has implications for the recombination behavior observed among mutants. Recombination analyses are most easily performed using conditional-lethal type. The cells are mixedly infected with two mutants at permissive conditions and then the nonpermissive conditions are applied to select for recombinants. The so called two-factor crosses are performed through simple pairwise infections. For single-segmented DNA viruses this approach allows the mutants to be ordered into complementation groups and the relative positions of mutations to be placed on a linear map. For segmented genome DNA viruses, however, due to the reassortant factors, the data obtained in two-factor crosses do not allow mutants to be ordered on to a linear map. Three-factor crosses involve the use of three mutations, with one of them being kept unselected during selection of recombinants between the other two. These crosses are useful for determining linkage relationships between mutants and for establishing the order of marker mutations. Due to reassortment the three-factor crosses are of less use in segmented viruses.

Recombination in DNA viruses of eucaryotes was observed for both animal and plant viruses. The recombination frequencies among pairs of temperature-sensitive mutants (two-factor crossings) were studied in herpes simplex virus. A linear dependence upon the distance between mutations and the fre-

Table 1 Functional analogy between phage and *E. coli* recombination proteins.

Phage	*E. coli*
T7 gene 3	RuvC, RecG
T4 UvsX	RecA
T4 gene 49	RuvC, RecG
T4 genes 46 and 47	RecBCD
I ORF in nin region	RecO, RecR, RecF
Rac recE gene	RecJ, RecQ

ORF, open reading frame.

quency was observed, suggesting the lack of specific signal sequences responsible for crossover events in this virus.

Three-factor crosses were performed using herpes simplex virus. Here, a syncytial plaque morphology mutation was used as an unselected marker for ts mutations. In the case of adenoviruses the host range determined by the helper function of two mutations has been used as a third marker between ts mutants.

Epstein–Barr virus (EBV), a member of the gamma herpesvirus family, is a DNA virus with a long double-stranded DNA genome which shows a high degree of variation among strains. This variation takes the form of single base changes, restriction site polymorphism, insertions and deletions. It was found that some EBV variants arose by DNA recombination events.

Homologous genetic recombination was observed in vaccinia virus (VV) and other poxviruses. This is evident from the high frequency of intertypic crossovers, the ease of the marker rescue and the isolation of viral recombinants. Recombination of VV DNA occurs both intra- and intermolecularly, and is dependent on DNA target size. A function of the viral DNA polymerase or viral DNA replication itself has been implicated and viral proteins with DNA strand transfer activity have been identified.

Intertypic crosses between ts mutants were selected for adenoviruses. Analysis of the segregation patterns of DNA fragments and their restriction enzyme polymorphism has allowed the specific ts mutations to be mapped on adenovirus genomes.

General recombination in somatic cells was observed for SV40 (a papovavirus). The authors tested recombination events from artificially constructed recombinant circular oligomers. While this type of recombination was high, homologous recombination in this type of DNA tumor viruses was rare.

The geminiviruses are a unique group of single-stranded plant DNA viruses. Intermolecular recombination between geminivirus DNAs has frequently been observed using various combinations of mutants. Likewise, intramolecular homologous recombination between tandem repeats of a geminivirus genome was found in agro-infected tobacco plants. Both homologous recombination model and replication-based recombination model were proposed to explain the observed events. In addition, deletions, insertions and other rearrangements have frequently been detected in geminivirus infections. These illegitimate recombination processes may rely on aberrant breakage–fusion events as well as errors of DNA replication, and may be inter- or intramolecular in nature.

Cauliflower mosaic virus (CaMV) is a plant DNA virus. It belongs to the pararetroviruses, which replicate through a reverse transcription step. A high recombination rate was observed *in vivo* for CaMV. The replication cycle of CaMV offers a variety of possibilities for recombination: recombination at the DNA level, which occurs in the nucleus, and at the RNA level, which occurs during reverse transcription in the cytoplasm. In general, it is difficult to decide by which recombination route a CaMV recombinant is obtained. However, such features as recombinational hot spots and apparent mismatch repair might be indicative. Namely, the presence of hot spots reflects replicative (RNA) recombination, while mismatch repair can occur during formation of heteroduplex intermediates and is thus indicative of DNA recombination. Recombination between CaMV strains and CaMV transgene mRNA was observed. It is believed that this type of recombination represents an RNA–RNA recombination event (during reverse transcription).

Recombination in RNA Viruses

RNA viruses utilize RNA as their genetic material. The potential for variation of the RNA genome is very large owing to a high mutation rate (during copying by RNA-dependent RNA polymerase) and to recombination. The terms of classic population biology do not describe RNA viruses. Instead, a term of quasispecies was proposed to reflect the nature of RNA virus populations. The processes of genetic recombination in plus-stranded RNA viruses probably do occur at the RNA level, as these viruses most likely do not go through DNA steps in their replication cycles. RNA recombination processes are generally categorized as either homologous or nonhomologous. In 1992, Lai postulated three classes of RNA recombination: homologous, aberrant homologous and nonhomologous. Homologous recombination occurs between two related RNA molecules at corresponding sites, although homologous RNA

recombination can also occur within a common region shared by otherwise unrelated RNA sequences. Aberrant homologous recombination involves crossovers between related RNAs, but does not occur at corresponding sites, leading to sequence insertions or deletions. Nonhomologous recombination occurs between unrelated RNA molecules. Slightly different definitions, based on mechanistic models and considerations, were proposed recently by several authors.

Although genetic recombination in RNA viruses such as influenza virus and poliovirus has been described, it has not been found in Newcastle disease virus. The complete nucleotide sequences of the genomic RNAs of a large number of RNA viruses belonging to different virus groups have been obtained. This revealed the relatedness of various animal, plant and other RNA viruses and allowed the definition of sequence rearrangements in the viral RNA genome.

Sequence rearrangements were found in the following animal plus-strand RNA viruses: picornaviruses: poliovirus and foot-and-mouth-disease virus (FMDV); in coronaviruses: mouse hepatitis coronavirus (MHV); Sindbis alphavirus (SIN); flock house nodavirus (FHV); and in bacteriophages $Q\beta$, and MS-2. Recombinants were found in bunyaviruses. Genetic rearrangements were also observed in other types of RNA viruses, including influenza virus, a minus-strand RNA virus, in retroviruses, and in double-stranded $\Phi 6$ bacteriophage. The following genomes of plant RNA viruses reveal RNA rearrangements: alfalfa mosaic virus (AlMV), beet necrotic yellow vein virus (BNYVV), bromoviruses (see below), hordeiviruses, luteoviruses, nepoviruses, tobamoviruses, tobraviruses, tombusviruses and turnip crinkle carmovirus (TCV).

Recombination by reassortment was observed for multisegmented animal RNA viruses, including influenza virus and double-stranded reoviruses and orbiviruses. The reassortment mechanism functioning in reoviruses is reflected by the fact that the interpretation of two-factor crosses (using for instance temperature-sensitive mutations) appears to be difficult: the mutants cannot be ordered on to a self-consistent linear map and quite often no linkage between mutants could be detected.

The examples in which host-derived sequences have recombined with viral RNAs include an ubiquitin-coding sequence of bovine diarrhea virus, a sequence from 28S rRNA inserted in the hemagglutinin gene of an influenza virus, and a tRNA sequence in Sindbis virus RNA. For plant viruses, several potato leafroll virus isolates contain sequences homologous to an exon of tobacco chloroplast. Acquisition of chloroplast sequences during RNA recombination was observed for brome mosaic virus.

The use of transgenic plants expressing viral RNA sequences has confirmed that plant RNA viruses are able to recombine with host mRNAs. This was shown for cowpea chlorotic mottle bromovirus (CCMV), red clover necrotic mosaic virus (RCNMV), potato virus Y potexvirus (PVY) and plum pox potyvirus (PPV).

Studies on the molecular mechanism of RNA recombination have progressed when experimental systems that supported the high frequency of crossovers were established. The available data on rearrangements in picornaviruses suggest a mechanism of template switching that occurs during minus-strand RNA synthesis. These rearrangements may be facilitated by the existence of identical or completely dissimilar signal sequences between the recombining RNA substrates.

Certain RNA viruses can produce both homologous and nonhomologous RNA recombinants. The molecular mechanism involved in the formation of homologous and nonhomologous recombinants was tested using an efficient recombination system of brome mosaic virus (BMV). A partially debilitating BMV RNA3 mutant was repaired *in vivo* by exchanges with the sequences of other BMV RNA components. Low recombination frequency was overcome by construction of RNA3-based recombination vectors, where recombinationally active sequences could be inserted and analyzed. It appeared that short base-paired regions between the two BMV RNA recombination substrates can target efficient nonhomologous recombination crossovers. A model invokes the formation of local RNA–RNA heteroduplexes to be responsible for targeting the RNA crossovers as a result of: (1) bringing the RNA substrates into a close proximity; and (2) slowing down or stalling the approaching replicase enzyme complex (**Fig. 3**).

Similarly, homologous RNA recombination was studied by inserting a BMV RNA2-derived sequence into the recombination vector. Both precise and imprecise crossovers were observed. Other RNA sequences revealed that the frequency of RNA2–RNA3 homologous crossovers depends upon sequence composition and tends to occur at hot spot regions that contain stretches of GC-rich alternating with AU-rich sequences. Such nucleotide composition may act as recombination activators during switching between RNA templates by the RNA replicase enzyme (**Fig. 4**). Overall, the data on BMV RNA recombination suggest that molecular mechanisms involved in the two types of crossovers in BMV differ from each other.

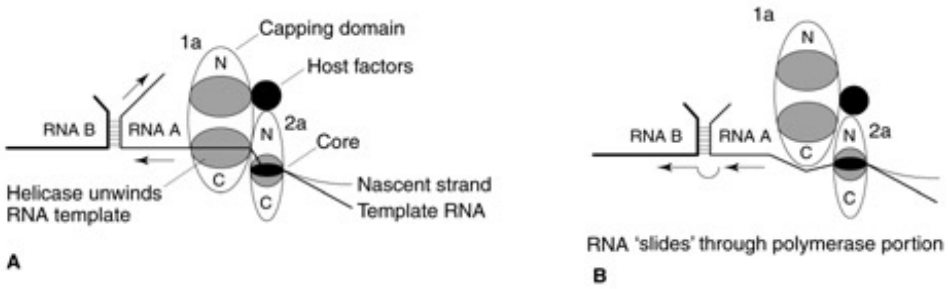

Figure 3 The model of strand switching by BMV replicase. The replicase is composed of the interacting host (smaller black circles) and viral (represented by ellipses) proteins. Functional domains of nucleotidyl transferase (capping enzyme), helicase and core RNA polymerase on proteins 1a and 2a are represented by smaller shaded ellipses. At the replication mode (**A**) the enzyme copies the original template through the double-stranded region (represented by a short 'ladder') due to the helicase action. During recombinational mode (**B**), the polymerase 'slides' under the double-stranded regions and changes the templates from RNA A to RNA B. The arrows indicate the direction of replicase migration. (Reproduced from Bujarski and Nagy (1994) Genetic RNA–RNA recombination in positive-stranded RNA viruses of plants. In: Paszkowski J (ed.) *Homologous Recombination and Gene Silencing in Plants.* Kluwer Academic.)

The term 'recombinosome' was proposed to describe a complex between the recombining RNAs, the replicase proteins and other (putative) factors involved in template switching events. The participation of replicase proteins of BMV in recombination was studied using a temperature-sensitive 1a of BMV protein mutant. This revealed a *5′* shift in crossover sites within the RNA1–RNA3 heteroduplex, suggesting that the helicase domain of 1a participates in heteroduplex-mediated crossovers. Likewise, a single amino acid mutation within the core domain of 2a protein and mutations within the N-terminal portion of 2a, the polymerase component of the replicase, inhibited the frequency of nonhomologous recombination in BMV. These studies confirm the participation of replicase proteins in recombination.

The role of replicase enzyme in RNA recombination was also studied in TCV, a small single component RNA virus that is associated with a number of subviral RNAs, such as satellite RNA D and chimeric RNA C. High frequency recombination was observed *in vivo* between RNAs C and D. A template switching model was proposed where viral replicase utilizes the nascent plus-strand of RNA D to reinitiate RNA elongation at a hairpin structure on the acceptor minus-strand RNA C template (**Fig. 5**).

The participation of TCV replicase in RNA recombination was studied *in vitro* with a TCV replicase preparation and a chimeric RNA template containing the *in vivo* hot spot region from RNA D joined to the hot spot region from RNA C. This demonstrated roles for a priming stem sequence in the RNA C portion and the TCV RNA-dependent RNA polymerase (RDRP) binding hairpin, also from the RNA3 portion. It probably reflects such late steps of the *in vivo* RNA recombination as strand transfer and primer elongation.

For coronaviruses, the animal RNA viruses containing a large RNA genome, recombination has been demonstrated between coronavirus genomes and defective-interfering RNAs, and it was postulated to account for the diversity in the genomic structure of these viruses. The mechanistic considerations suggest the nonprocessive nature of the coronavirus RNA polymerase, which might be responsible for recombination. Similarly, RNA recombination in nodaviruses, two-partite RNA viruses, occurs between RNA segments at a site, where the nascent strand could form a base-paired region with the acceptor template. Such factors as template secondary structure and the similarity of the crossover sites to an origin of replication seem to influence the choice of recombination site. A model of recombination where the polymerase interacts directly with the acceptor nodavirus RNA was postulated.

A copy-choice template switching mechanism was also suggested for recombination in a double-stranded bacteriophage Φ6. Here, the crossovers occur inside the virus capsid structure. Apparently, the crossovers can occur in regions that share little sequence similarity and the frequency of recombinants can be enhanced by conditions that prevent the minus-strand synthesis.

The bacteriophage Qβ has emerged as a unique RNA virus system for the study of RNA recombination both *in vivo* and in cell-free systems. It was demonstrated for the first time in this virus that RNA recombination can occur not by polymerase template switching events but rather via a splicing-type RNA recombination mechanism.

Genetic RNA recombination has been observed in retroviruses. Here, the efficient recombinant jumpings are secured by reverse transcriptase. In fact, the retrovirus system represents a well-established model

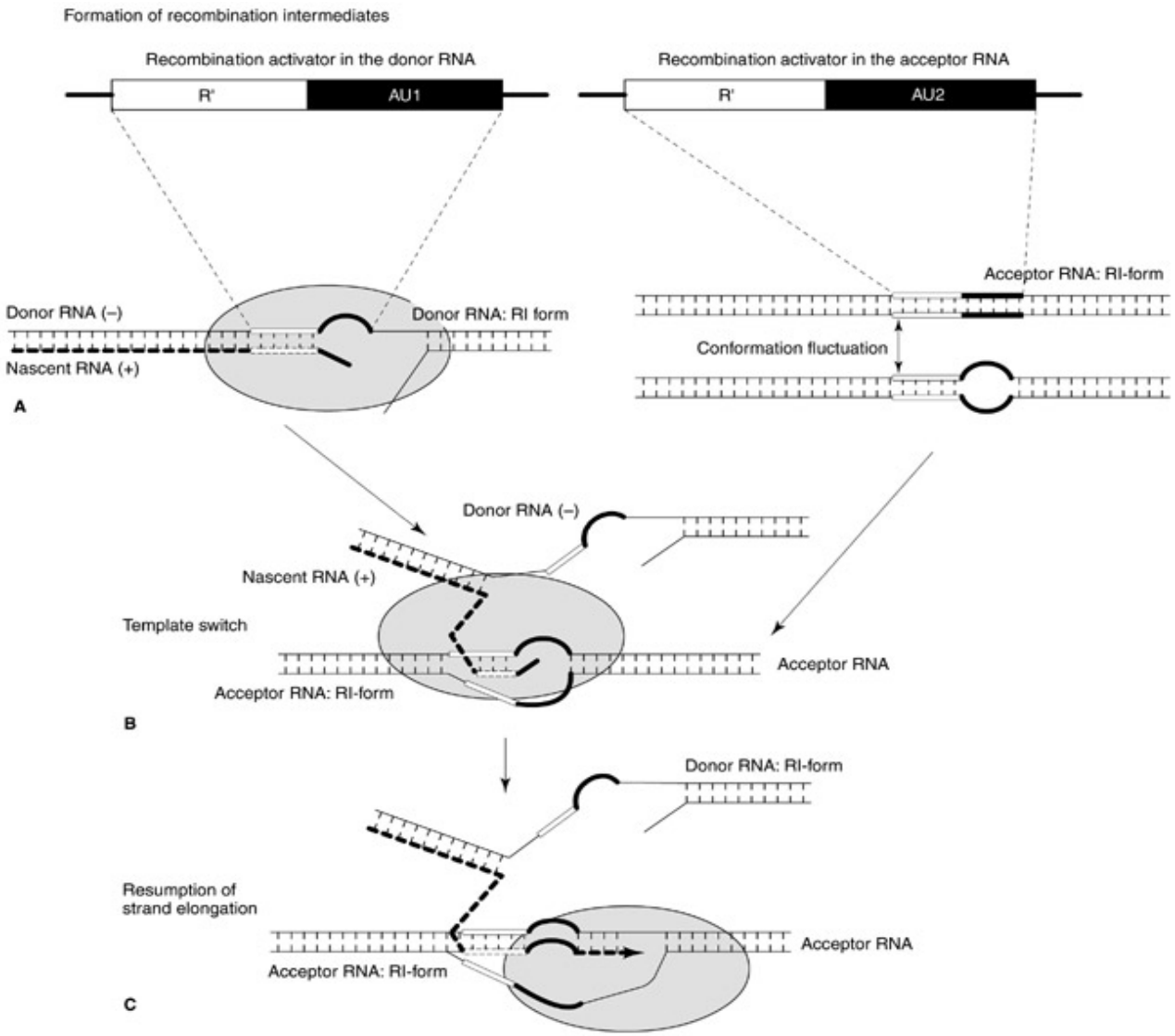

Figure 4 Processive template switching model explaining the formation of homologous recombination hot spots within the recombination activator sequences. (**A**) Template switching of the BMV replicase occurs during positive-strand synthesis. Although localized double-stranded replicative intermediates (RIs) are shown, the existence of single-stranded RNAs with negative polarity is also possible (not shown). The weak base pairing within the AU-rich region (shown on the left side of the diagram with a black line) can facilitate the release of the 3′ end of the incomplete nascent RNA. The weak base pairing within the AU-rich region can also facilitate the temporary formation of a bubble structure in the RI of the acceptor strand (gray line on the right side). The replicase (large shadowed ellipse) pauses on the donor strand at the UA-rich region, and the very 3′ end of the nascent strand disengages from the original template strand. (**B**) The released 3′ end of the nascent strand hybridizes to the acceptor strand facilitated by the bubble structure. Hybridization of the upstream located R′ (shown by empty boxes) stabilizes the recombination intermediate. (**C**) The viral replicase resumes chain elongation on the acceptor strand (shown by an arrowhead). This leads to the formation of homologous recombinant RNA3s. (Reproduced with permission from a paper by Nagy and Bujarski (1997) *J. Virol.* 71(8): 3808.)

of the polymerase/template switching reactions both *in vivo* and *in vitro*. Apparently, the virally encoded reverse transcriptases are evolutionarily selected to secure jumping during reverse transcription reactions. The recombinant jumpings are responsible for both inter- and intramolecular template switching, and for

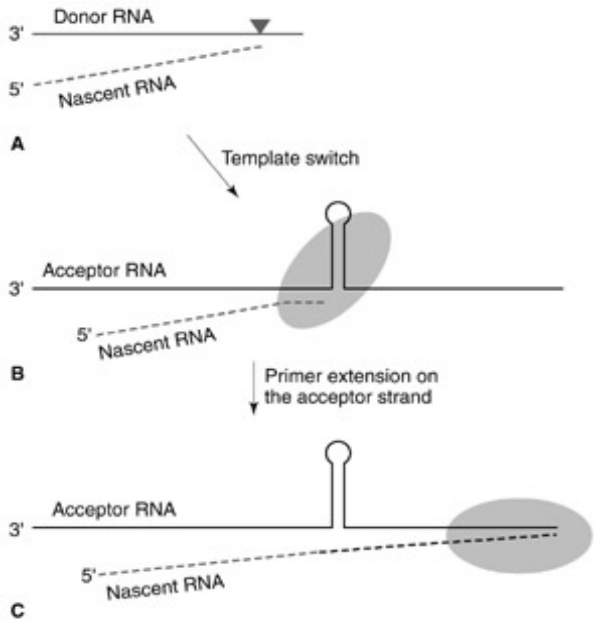

Figure 5 Model for RNA recombination in the turnip crinkle virus (TCV) system. (**A**) The TCV RDRP copying minus-strand sat-RNA D either reaches the natural 5′ end or pauses at some positions (mainly 13 nucleotides from the 5′ end, as indicated by a triangle) likely due to the presence of a protein-binding site. (**B**) The TCV RDRP, which is still associated with the nascent sat-RNA D strand, switches to the acceptor template (minus-strand sat-RNA C or TCV) facilitated by either the motif I or motif III hairpins. Hybridization between the nascent strand and the acceptor strand may stabilize the recombination intermediate. (**C**) The TCV RDRP reinitiates RNA synthesis using the 3′ end of the nascent sat-RNA D as a promoter. Further copying of the acceptor RNA by the RDRP results in a recombinant RNA molecule. (Reproduced with permission from Simon *et al* (1996) *Semin. Virol.* 7: 373.)

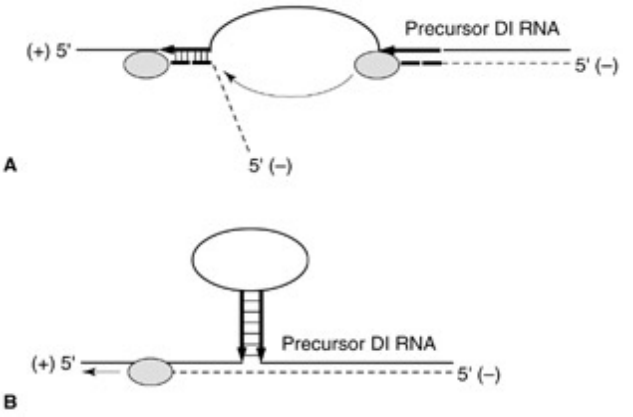

Figure 6 Proposed replicase-mediated model for deletion of segments from precursor DI RNAs. RNA templates are depicted as solid lines while nascent strands are shown as dashed lines. The viral replicase is represented by shaded ovals and its path is indicated by thin arrows. Deletion of segments from precursor DI RNA molecules may be facilitated by tracts (bold arrows) of sequence (**A**) identity or (**B**) complementarity, as discussed in the text. (Reproduced with permission from White A (1996) Formation and evolution of Tombusvirus defective interfering RNAs. *Semin. Virol. 7:* 409.)

the formation of defective retroviral genomes. They contribute significantly to genetic variability of retroviruses.

Defective-Interfering RNAs

Defective-interfering (DI) RNAs are subviral RNA molecules derived from the helper virus genomic RNA and typically interfere with helper virus accumulation and affect symptoms produced by the helper virus. Paul von Magnus was first (in 1954) to report DI RNAs in influenza virus. Later, DI RNAs were observed in a majority of animal and in many plant RNA virus infections. Naturally-occurring DI RNAs have been identified during infection with several coronavirus species. These molecules appear to arise by a polymerase strand-switching mechanism. In fact the DI RNAs were used in coronavirus research, to study the mechanism of high-frequency, site-specific RNA recombination events that progress through leader acquisition during RNA replication, and as vehicles for the generation of targeted recombinants of the parental virus genome.

For plant tombusviruses and carmoviruses the DI RNAs reveal a consistent pattern of rearranged mosaic-type sequences flanked by unmodified terminal regions. Analysis of these DI RNAs suggests that in some cases the base-pairing between an incomplete replicase-associated nascent strand and the acceptor template can mediate the selection of the rearrangement sites (**Fig. 6**).

Single deletion-type DI RNAs were isolated from infections with several viruses, including beet necrotic yellow vein furovirus (BNYVV), soil-borne wheat mosaic furovirus (SBWMV), peanut clump furovirus (PCV), clover yellow mosaic potexvirus (CYMV), sonchus yellow net rhabdovirus (SYNRV), and tomato spotted wilt tospovirus (TSWV). Such factors as the length of the defective RNA or its coding capacity seem to affect the selection of DI RNAs during infection.

For broad bean mottle bromovirus (BBMV) the naturally existing DI RNAs were found to be derived by single deletions in the RNA2 component. A secondary structure-mediated model for BBMV DI RNAs, where local complementary regions bring the remote parts of RNA2 together to facilitate the crossover events, has been proposed (**Fig. 5**). Similar to BBMV, single deletion-type DI RNAs have been found associated with cucumoviruses. Overall, it